

MULTI-RATE TRANSCODER FOR DIGITAL STREAMS

BACKGROUND OF THE INVENTION:

5 The present invention relates to the multi-rate transcoding of data signals such as video, audio and/or multimedia, and is particularly applicable to the streaming of data from a server to a client over a network, such as the Internet or a satellite or cable television system.

10 Although the invention is described herein in connection with a multi-rate transcoder for compressed video applications, it should be appreciated that the inventive concepts are applicable to any type of digital data stream in which there is a need to provide the same data at multiple rates. Accordingly, the invention and the appended claims are not limited to the video applications specifically disclosed herein.

15 Most network applications can be divided into two main components. These are a client and a server. There are numerous examples of clients and servers. In particular, streaming involves sending movies or other content, such as audio and/or other multimedia formats from a server to a client over a network, such as Internet. Streaming is different from simple file transfer, in that the client plays the movie (or other content) as it arrives from the network, rather than waiting for the entire movie to be received before playing. Real-time streams can be sent one-to-one (unicast) or one-to-many (multicast). In unicast, the client contacts the server to request a movie. The server then
20 replies to the client with the information about the requested movie. The actual movie streams are then sent to the client. In multicast, one copy of each stream of a movie is "broadcast" over each branch of a network. Clients receive the streams by "joining" the multicast. Not all routers support multicasting.

25 A client, such as a cable or satellite television subscriber, normally communicates with one server at a time. For example, the client terminal sends a request to a server (unicast) and receives streams from a server (both unicast and multicast). From the server's perspective at any given time, it is common for the server to be communicating

with multiple clients. For example, in unicast, a server may send out streams to different clients simultaneously, and in multicast, multiple clients receive streams broadcast by a server.

With recent advances in video compression, the size of raw video material can be significantly reduced by employing compression technology, such as, MPEG 1, 2, 4 (where MPEG are standards promulgated by the Moving Picture Experts Group) and ITU-T H.261, 3, 3+, or L (where ITU standards are promulgated by the International Telecommunications Union), without sacrificing picture quality. Hence, it is expected that there will be a greater amount of pre-compressed video material available at the server end in the future. A server may therefore need to have the capability to adjust the rates of pre-compressed video material stored at the server end in order to accommodate the bandwidth requirements of different clients. A transcoder is one options for addressing this problem. By definition, a transcoder is a device for converting a pre-compressed bit stream into another bit stream with different formats. Such different formats can include resolution, bit rates, and other variable parameters. The present invention provides a transcoder that provides rate conversion of a single input stream to provide a plurality of different rate output streams.

When client (e.g., a set-top box for receiving television services at a subscriber's home) requests a movie, for example, it often provides information about the network condition at the client end as well. The server can then transcode the pre-compressed streams of the requested movie at a bandwidth suitable for the client end. An issue that has to be addressed, however, is that the server may receive multiple requests for the same movies from different clients. Moreover, the bandwidth at the client ends can be very different. For example, a phone modem at 56Kb/s and a Cable modem at a few Mb/s have different abilities to receive data from the server. Hence, it would be advantageous for a single compressed video bit stream at the server end be transcoded into several different new streams at different rates. A straightforward solution would be to use different transcoders to transcode the same input video stream at different rates in response to the client requests. This solution is not practical, however, as it requires the

server to have as many transcoders as the clients it serves, which would be cost prohibitive.

Another solution would be for one transcoder at the server end to process the same video stream as many times as requested by different clients, but one at a time. That is, before the server completes the request by one client, the rest of the clients will have to wait. Thus, to address the issue of one server and multiple services, either the system operator will have to install more transcoders or the clients will have to wait for their turn to receive the requested service (e.g., movie). Moreover, by the above two solutions, all the function blocks in a transcoder are required to be implemented for each request for the same video bit stream.

It would be advantageous to provide a reliable and cost effective transcoding solution that enables concurrent delivery of a service to a population of users (clients) at different bit rates, depending on the client requirements. The present invention provides a multi-rate transcoding solution having the aforementioned and other advantages.

SUMMARY OF THE INVENTION

In accordance with the invention, transcoder apparatus is provided for an encoded bit stream. A data processor extracts overhead data from the bit stream. A decoder is provided for at least partially decoding the bit stream. A rate control processor re-encodes the at least partially decoded bit stream at different rates, to produce multiple re-encoded bit streams having different rates. A multiplexer combines the overhead data with each re-encoded bit stream, thereby providing multiple versions of the encoded bit stream at different rates.

In a preferred embodiment, a multiplexer provides the multiple versions of the encoded bit streams substantially simultaneously. The encoded bit stream can be a compressed video bit stream, with the transcoder being located at a streaming video server for providing multiple versions of the video bit stream to different clients, at different rates, substantially simultaneously.

In an implementation where the encoded bit stream is a compressed video bit stream, the overhead data can comprise at least one of video object sequence (VOS), video object (VO), video object layer (VOL), video object plane (VOP), group of video object planes (GOV) and motion vector (MV) data. The overhead data can, for example, be extracted from packet headers contained in the encoded bit stream. It should be appreciated that while certain of the terms used herein, such as VO, VOL, VOP, and GOV may be specific to a particular standard, such as MPEG-4, the concepts embodied thereby may be found in other standards. Thus, the use herein and in the claims of a term found in a particular standard is not meant to be limited to that standard, but is intended to be broadly construed as applicable to any standard or technique in accordance with the present invention.

In one embodiment, the rate control processor re-encodes said at least partially decoded bit stream a plurality of times to produce the multiple re-encoded bit streams on a sequential basis. More specifically, the rate control processor can re-encode the at least partially decoded bit stream separately for each of the multiple re-encoded bit streams.

Typically, the encoded bit stream is received at a first rate. The rate control processor can operate at a second rate of at least N times said first rate, where N is the number of re-encoded bit streams provided. In this manner, the re-encoded bit streams are all provided substantially concurrently with the original compressed video bit stream.

5 In an advantageous embodiment, first functions that do not effect the rates of the re-encoded bit streams are performed only once on the encoded bit stream. Second functions that effect the rates are performed separately for each re-encoded bit stream.

In an embodiment where the encoded bit stream is a compressed video bit stream, the first functions may comprise at least one of variable length decoding and dequantization. The second functions may comprise at least one of requantization, variable length coding, and motion compensation.

10 The rate control processor can comprise a plurality of encoders operating in parallel to produce the multiple re-encoded bit streams. The re-encoded bitstreams can be provided as variable bitrate streams and/or constant bitrate streams, depending on the client requirements.

15 In order to efficiently provide a plurality of different rate output streams, processor cycles of the rate control processor can be monitored, and at least one processing step can be skipped in the event the number of processing cycles available to complete a rate control operation would otherwise be insufficient. For example, where the encoded bit stream comprises a compressed video bit stream, at least one of a motion compensation step and a discrete cosine transformation (DCT) step can be skipped for a bi-directionally predicted (B) frame in the event the number of processing cycles available to complete a rate control operation would otherwise be insufficient.

20 In another optimization technique, the rate control processor can be implemented to re-encode the at least partially decoded bit stream a plurality of times to produce the multiple re-encoded bit streams on a sequential basis, while selectively skipping at least one processing step for fewer than all of the multiple re-encoded bit streams.

25 Corresponding methods are also provided.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram of a client server network;

Figure 2 illustrates the concept of multiple clients being served by the same server over a network such as the Internet;

5 Figure 3 is a block diagram of a prior art implementation wherein a plurality of transcoders are used to provide a like plurality of output streams from a compressed video bit stream;

Figure 4 is a block diagram of a prior art implementation in which one transcoder is used to process the same compressed video bit stream in a sequential manner;

10 Figure 5 is a block diagram illustrating the present invention, wherein multiple streams at different rates are produced simultaneously from a compressed video bit stream provided as input; and

15 Figure 6 is a block diagram of an embodiment where a macroblock is decoded only once, and re-encoded by different encoders to provide a plurality of streams at different rates.

DETAILED DESCRIPTION OF THE INVENTION

According to the present invention, transcoder apparatus is provided for simultaneously providing a plurality of bit streams encoded at different rates from a single input bit stream, such as a compressed video bit stream.

5 A basic concept of the invention is the provision of a multi-rate transcoder. This transcoder can take a single compressed video bit stream as input and substantially simultaneously generate multiple streams at different rates at the request of different clients.

10 Figure 1 illustrates a simple network application, in which a client 12 communicates with a server via communication path 14. Such client-server networks are well known in the art. While a client normally communicates with one server at a time, typically sending a request to the server and receiving an information data stream in response, a server will typically communicate with a plurality of clients simultaneously. This is illustrated in Figure 2, wherein server 10 communicates with multiple clients 12A, 12B ... 12C via, 15 e.g., the Internet 16. The clients can be any of a number of different appliances, including, without limitation, personal computers (PCs), wireless Internet appliances, television set-top boxes, point of sale terminals, wagering terminals, consumer appliances, photocopiers, printers, industrial equipment, automotive systems, and the like. Typically, the different clients (even those of the same general type, such as set-top 20 boxes) will have different processing capabilities and requirements for receiving data from the server. Such data is often communicated from the server to the client in the form of packetized data streams that comply with an industry wide standard. An example of one such standard is that promulgated by the Motion Picture Experts Group for the communication of compressed digital video signals, and known as MPEG-2. Other 25 standards exist and new standards are currently under development, and the invention is not limited to any such transmission standard.

In a television embodiment, which is one application for the present invention, a client (e.g., cable or satellite television subscriber) may request a movie. As part of the

request, the server may receive information about the network condition at the client end of the communication path, and can then transcode pre-compressed streams which carry the requested movie at the bandwidth of the client end. In the event the server receives multiple requests for the same movie from different clients, each client having a different bandwidth requirement, it will be necessary for the transcoder to provide multiple instances of the move at the different required bandwidths.

A straightforward solution to this problem is to transcode the same input video stream at different rates as required by the different clients, as illustrated in Figure 3. A disadvantage of this solution is that the server is required to have as many different transcoders as the clients it serves. This is not realistic.

Another solution is for the transcoder at the server to process the same video stream as many times as requested by different clients, but one at a time. In such an embodiment, before the server completes the request by one client, the rest of the clients will have to wait their turn to see the requested movie or other service requested. Such an embodiment is shown in Figure 4, wherein the transcoder 20 first provides Stream 1 from the compressed video bit stream, then provides Stream 2 after a delay 22, and then provides Stream N after another delay 22 (or series of delays depending on intervening requests), and so forth. Again, this is not a realistic solution as real-time provision of services on request is not provided.

The present invention overcomes the problems noted above by providing a multi-rate transcoder which processes a single compressed bit stream, such as a video bit stream, and concurrently generates multiple streams therefrom at different rates upon request from a plurality of clients. Fig. 5 shows one possible architecture for a multi-rate transcoder in accordance with the invention. Assume a pre-compressed video bit stream arrives. This stream is applied to a header processor 30 and a variable length decoder 32 (assuming that the bit stream contains variable length packets). At the header processor 30, header segments of the pre-compressed video bit stream, such as, video object sequence (VOS), video object (VO), video object layer (VOL), group of video object planes (GOV) and video object plane (VOP) will be scanned out only once for

subsequent multiplexing in a multiplexer 60 with the various different rate streams to be produced.

Next, at the macroblock (MB) level of video processing (MBs are defined, for example, in the widely available MPEG-2 specification), since motion vector (MV) fields are re-used, they can simply be copied into new streams. This is illustrated in block 30, which shows that the MV fields are treated like the header information described above.

For purposes of illustration, assume that N streams need to be generated at N different rates as requested by N clients. At the MB/block level, the shadowed function blocks, including quantizer (Q(i)) 42, dequantizer (DeQ(i)) 46, variable length coder (VLC) 44, Rate Control subsystem 36, Discrete Cosine Transform processor (DCT) 56 and inverse DCT (IDCT) 50, as well as motion compensation engine (MC) 54 will perform N times, one for each requested rate. As shown in Figure 5, the rate control function is controlled by a processor 40, which can be a microprocessor as well known in the art. An example of a rate control processor that can be implemented for use in connection with the present invention is disclosed in commonly assigned U.S. patent application 09/198,867 filed on November 28, 1998 and entitled "Rate Control for an MPEG Transcoder Without A Priori Knowledge of Picture Type."

The plain function blocks, such as variable length decoder (VLD) 32 and Dequantizer (DeQ) 38, are implemented only once in accordance with the present invention. Specifically, each input MB in compressed bits will go through the VLD 32 and DeQ 38. If the MB is in intraframe mode, it is passed via gate 40 to quantization block (Q(i)) 42 directly. If in interframe mode, the MB in the DCT domain is first motion compensated via gate 40, which receives quantization errors from the previous frame stored in Frame Buffers 52 after being processed by DeQ(i) 46, summer 48, and IDCT 50, and subsequently processed by Motion Compensator (MC) 54 and DCT 56 in a conventional manner. Rate control 36, in accordance with signals from processor 34, will assign the MB a new quantization factor, Q(i), for each new rate requested by client i. The quantized coefficient errors are variable length coded at 44, which generates a new stream at a new rate. In the feedback loop for the motion compensation stage, the

coefficient errors dequantized at 46 are subtracted from the coefficient errors at 48 and stored in Frame Buffers 52, one for each different rate stream. Note that the motion compensation uses the same motion vector (MV) field decoded from the VLD, but different reference (quantization error) frames for different new streams.

5 Various new features of the proposed multi-rate transcoder are as follows:

1. The multi-rate transcoder can be applied to almost all video coding standards, such as MPEG 1, 2, 4 and H.261, 3, 3+, L.
2. The principle of the proposed multi-rate transcoder is one decode operation and multiple encode operations. This principle can be used for almost all transcoder architectures, such as cascaded transcoders with or without reuse of MV fields, and transcoders with various shortcuts. For example, in Fig. 6, an input MB is decoded only once by a decoder 62, but, it is re-encoded by different encoders 64A, 64B, ...64N, resulting in different streams after being multiplexed in multiplexer 60 with the common header data from header processor 30.
3. The output streams can be either constant bit rate (CBR) or variable bit rate (VBR), depending upon the client requests.
4. Within a limited CPU cycle, the multi-rate transcoder may choose to abandon some of the function blocks of Figure 5. For example, if short of CPU cycles, one or more of the function blocks such as DCT 56, MC 54, Frame Buffers 52, IDCT 50, summer 48 and/or DeQ(i) 46 may not be performed for bi-directionally predicted (B)frames, which can save a significant number of CPU cycles. Since B frames are never used as reference frames, the overall video quality will not be affected too much and no error propagation will be caused. The multi-rate transcoder can also selectively stop performing some of the function blocks for different streams, depending upon the priorities.

Various advantages of the inventive multi-rate transcoder over prior art solutions are as follows:

1. A server is able to offer multiple new streams at different rates simultaneously.

2. Header segments (VOS, VO, VOL, GOV and VOP) are scanned only once and copied into different new streams directly.
3. Decoding function blocks, such as VLD and DeQ, only have to be implemented once for different streams at different rates.
- 5 4. The multi-rate transcoder can choose to perform either all or some of the function blocks for different streams.

It should now be appreciated that the present invention provides a multi-rate transcoder that provides multiple output streams at different rates from a single input stream. The output streams are provided concurrently, without the need to duplicate encoder structure that is already available. Common processing of the input stream is provided to the extent possible for all of the different output streams. Processing functions that are not common to all of the N output streams can be implemented at a rate that is essentially N times the rate it takes to process a single stream, so that all of the streams can be provided on a real-time basis. Hardware and software requirements are reduced as compared with prior art multiple transcoder solutions, and processing requirements are reduced as compared to a single transcoder solution.

Although the invention has been described in accordance with particular embodiments thereof, it should be appreciated that various adaptations and modifications may be made thereto without departing from the scope of the invention as set forth in the following claims.